

Estimação de Matrizes de Tráfego de Backbones utilizando Restrições de Proporcionalidade em Modelo de Programação Linear

Arnoldo N. da Silva¹, Marcelo H. O. Lima², Silas S. L. Pereira², Daniel N. Nobre²,
Jorge L. C. Silva², J. E. Bessa Maia², Paulo R. F. Cunha¹

¹Centro de Informática – Universidade Federal de Pernambuco (UFPE)
Caixa Postal 7851 – 50732-970 – Recife, PE – Brasil

²Departamento de Estatística e Computação - Universidade Estadual do Ceará (UECE)
Av. Paranajana, 1700 - 60740-903 – Fortaleza, CE – Brasil

{ans2,prfc}@cin.ufpe.br,
{marcelo,silas,danieln,jlcs,jmaia}@larces.uece.br

Abstract. *Traffic Matrices (TM) hold information about traffic flow between all node pairs in a network. Measurement data on traffic flows may have a high computational cost, making the use of TM estimation techniques a necessity. In this work, we propose a new set of constrains for the linear programming problem to limit the number of viable solutions, which represent the traffic flow between node pairs. The new set of constrains is based upon the network manager's expertise. Proportionality constrains were added to the model and, then, the obtained results were compared with previous results in the literature. We also compared the linear programming model with new constrains with other known techniques and it proved to be better, giving smaller estimate errors.*

Resumo. *Matrizes de tráfego (MT) contêm informações a respeito do fluxo de tráfego entre todos os possíveis pares de nós em uma rede. A medição de dados sobre fluxos de tráfego pode gerar um alto custo computacional. Em função disso, surge a necessidade do uso de técnicas de estimação de matriz de tráfego. Este trabalho propõe um novo conjunto de restrições para uso em programação linear que limita o conjunto de soluções viáveis para estimar medidas de tráfego entre pares de nós, proporcionando melhores soluções que as obtidas anteriormente. O conjunto de restrições é baseado em informações estruturais da rede. As restrições de proporcionalidade foram acrescentadas ao modelo e resultados obtidos foram comparados. Os experimentos mostram que, entre os modelos estudados, as novas restrições utilizadas tendem a minimizar os erros dos valores estimados.*

1. Introdução

Matrizes de tráfego (MT) refletem o volume de tráfego que flui entre todos os possíveis pares origem e destino na rede. A aquisição de dados medidos para preencher uma matriz de tráfego apresenta alto custo de coleta e computacional [Medina et al 2002]. Isto torna necessário o uso de técnicas de estimação baseadas em informações parciais. A estimação de matrizes de tráfego é fundamental para muitas questões da engenharia de tráfego, auxiliando no planejamento do crescimento de redes e diagnósticos de problemas.

Várias técnicas já foram implementadas para a resolução do problema de estimação de matriz de tráfego. [Medina *et al.* 2002] apresenta soluções baseadas nas técnicas utilizando Redes Bayseanas [Tebaldi *et al.* 1998], Maximização de Expectativa [Cao 2000] e Programação Linear [Goldschmidt 2000], comparando os resultados em função do erro de estimação. Outras abordagens recentes tratam este problema, sendo que uma delas utiliza a teoria da informação [Zhang *et al.* 2003a]. Teixeira *et al.* (2005) avalia o efeito de mudanças no roteamento na variação das matrizes de tráfego. Em [Soule *et al.* 2005] é feita uma avaliação do equilíbrio entre a adoção de medições, o uso de inferência e modelagem para a estimação de matrizes de tráfego. [Medina *et al.* 2002] utilizou Programação Linear (PL) com restrições relacionadas com os enlaces (links), conservação de fluxos e com a positividade dos valores dos fluxos, sendo que os resultados foram insatisfatórios. Desta forma, este trabalho propõe a inclusão de novas restrições provenientes do conhecimento do administrador da rede com o objetivo de obter melhores resultados com a aplicação de PL.

A abordagem considera um ambiente de *backbone* de internet. Os dados medidos a serem usados para a estimação equivalem à quantidade de bytes ou pacotes de entrada em cada nó coletado via SNMP (Simple Network Management Protocol). Para a validação do trabalho foram utilizados modelos sintéticos de dados e medidas obtidas a partir de gerador randômico de tráfego. A implementação computacional foi modelada e testada no Matlab [MATLAB 1991].

2. Estimação de Matrizes de Tráfego utilizando Programação Linear

Um fluxo Origem-Destino (OD) consiste de todo tráfego que entra na rede em um ponto de ingresso comum e sai em um outro ponto da rede. O tráfego observado em um enlace da rede é formado pela superposição de fluxos OD. Seja c o número de pares OD de uma rede formada por n nós, então $c = n \times (n - 1)$. Os fluxos ODs estão relacionados com os enlaces através da seguinte equação linear:

$$Y = AX \quad (1)$$

Onde, $Y = (y_1, y_2, \dots, y_r)$ é o vetor que representa as medidas do tráfego nos enlaces, sendo r o número de enlaces. X é um vetor de pares ordenados que representa o tráfego dos fluxos ODs. A é a matriz de roteamento do tipo $\{0, 1\}$ e tamanho $(\# \text{ enlaces}) \times (\# \text{ fluxos})$, onde $A_{ij} = 1$ se o fluxo j passa pelo enlace i e $A_{ij} = 0$ caso contrário.

O problema principal consiste em estimar X , ou seja, encontrar um conjunto de fluxos OD que poderá reproduzir Y o mais próximo possível do real. Considerando que nas redes o número de pares OD é muito maior que o número de enlaces na rede, Obtém-se um número infinito de soluções possíveis para o vetor X . Devido o problema de estimação de matriz de tráfego exigir relacionamentos lineares, o mesmo pode ser formulado como um modelo de programação linear. Sabendo que o tráfego em cada enlace é formado pela soma dos fluxos OD que utilizam esse enlace, uma função objetivo que natural para este problema de otimização é:

$$\max \sum_{j=1}^c w_j X_j, \quad (2)$$

onde w_j é um peso para o par OD j e implica em maximizar a carga de tráfego que a rede suporta. Duas estratégias de definição dos pesos frequentemente utilizadas na

literatura são: (i) pesos unitários, os quais não atribuem qualquer prioridade entre os fluxos, e (ii) pesos baseados na distância da origem ao destino medida em número de saltos. A função objetivo é sujeita as condições dos enlaces:

$$\sum_{j=1}^c A_{lj} X_j \leq Y_l \quad l=1, \dots, r \quad (3)$$

A equação (3), originária da equação (1), mostra que a soma dos fluxos OD que passam pelo enlace Y_l não pode ser maior que o valor do tráfego medido naquele enlace.

As restrições de conservação de fluxo são dadas por [Goldschmidt 2000]:

$$\sum y_e - \sum y_s = \sum x_d - \sum x_o, \quad (4)$$

onde Y_e representa a quantidade de tráfego de entrada de cada enlace e Y_s a quantidade de tráfego de saída de cada enlace. X_d representa a quantidade de tráfego destinado ao nó e X_o a quantidade de tráfego originada neste nó. As restrições de positividade são representadas por $X_j \geq 0, \forall j$. Neste trabalho um novo conjunto de restrições, denominado de proporcionalidade, foi adicionado ao sistema. Estas restrições são baseadas na proporção da carga de tráfego em cada um dos nós relativa ao total da carga da rede. O administrador estima estes percentuais através do histórico da rede.

$$\sum_j X_{ij} - \lambda_i \sum_i \sum_j X_{ij} = 0, \quad i=1, \dots, n \quad (5)$$

$$\sum_i X_{ij} - \lambda_j \sum_i \sum_j X_{ij} = 0, \quad j=1, \dots, n \quad (6)$$

As equações (5) e (6) mostram, respectivamente, que o somatório do tráfego criado em um nó e o somatório do tráfego destinado a um nó é, cada um deles, igual ao somatório de todos os fluxos OD da rede multiplicado pelo fator de proporcionalidade λ .

2.1. Rede Estudada

A topologia da rede (Figura 1a) estudada é baseada em [Medina et al., 2002] onde os nós representam os PoPs. Os valores nas conexões dos nós representam as quantidades de tráfego nos enlaces. O objetivo é estimar a quantidade de tráfego em cada fluxo OD, dado a quantidade de tráfego total coletado na entrada e saída de cada nó. Os valores reais (Figura 2b) de fluxos OD para cada par serão comparados com os estimados.

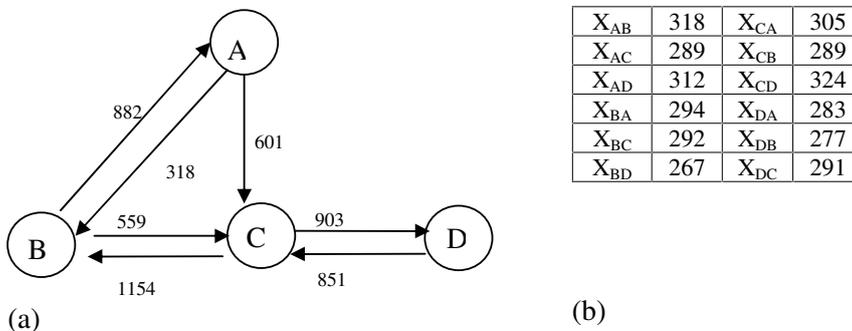


Figura 1: (a) Topologia usada como estudo de caso; (b) Valores reais do tráfego para cada par de fluxo OD.

A topologia em estudo apresenta pares de nós com enlaces próprios, com ligação direta entre os nós, e pares onde o tráfego precisa ser roteado via nós intermediários. A equação $AX = Y$ compõe, para o caso em estudo, o sistema linear abaixo:

$$\begin{bmatrix} 318 \\ 601 \\ 882 \\ 559 \\ 1154 \\ 903 \\ 851 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} X_{AB} \\ X_{AC} \\ X_{AD} \\ X_{BA} \\ X_{BC} \\ X_{BD} \\ X_{CA} \\ X_{CB} \\ X_{CD} \\ X_{DA} \\ X_{DB} \\ X_{DC} \end{bmatrix}$$

O sistema linear gerado pelo problema foi resolvido neste trabalho através da PL. Foram testados 3 casos variando as restrições e a função objetivo.

2.2 Estudo de Casos

Para cada um dos 3 casos foram definidos modelos de sistemas lineares, os quais foram executados no Matlab utilizando a função *linprog*.

2.2.1. Caso 1

Este primeiro estudo de caso é o mesmo modelo estudado em [Medina et al 2002]. A função objetivo tem pesos 1 ($w_i = 1$).

$$\max X_{AB} + X_{AC} + X_{AD} + X_{BA} + X_{BC} + X_{BD} + X_{CA} + X_{CB} + X_{CD} + X_{DA} + X_{DB} + X_{DC} \quad (7)$$

Aplicando as equações (3) e (4), temos respectivamente as restrições de link (Figura 2a) e as restrições de conservação de fluxo (Figura 2b):

$ \begin{aligned} X_{AB} &\leq 318 \\ X_{AB} + X_{AC} &\leq 601 \\ X_{BA} + X_{CA} + X_{DA} &\leq 882 \\ X_{BC} + X_{BD} &\leq 559 \\ X_{CA} + X_{CB} + X_{DA} + X_{DB} &\leq 1154 \\ X_{AC} + X_{BD} + X_{CD} &\leq 903 \\ X_{DA} + X_{DB} + X_{DC} &\leq 851 \end{aligned} $ <p>(a)</p>	$ \begin{aligned} X_{AB} + X_{AC} + X_{AD} - X_{BA} - X_{CA} - X_{DA} &= 37 \\ -X_{AB} + X_{BA} + X_{BC} + X_{BD} - X_{CB} - X_{DB} &= -31 \\ -X_{AC} - X_{BC} + X_{CA} + X_{CB} + X_{CD} - X_{DC} &= 46 \\ -X_{AD} - X_{BD} - X_{CD} + X_{CA} + X_{CB} + X_{CD} &= -52 \end{aligned} $ <p>(b)</p>
---	---

Figura 2: (a) Restrições de link; (b) Restrições de conservação de fluxo.

2.2.2 Caso 2

No segundo estudo de caso foram mantidas a mesma função objetivo e as restrições de conservação de fluxo do Caso 1. Porém, foram adicionadas as restrições de proporcionalidade. Como temos um modelo de topologia sintético, consideramos tais informações a partir dos resultados reais dos valores de fluxos OD para calcular a

proporção de carga em cada nó. Ou seja, como pode ser visto nas equações (8), (9), (10) e (11) o nós A e C participam com 26% da carga total, enquanto que os nós B e D proporcionam 24% da carga total.

$$X_{AB} + X_{AC} + X_{AD} = 0.26 \sum X \quad (8)$$

$$X_{BA} + X_{BC} + X_{BD} = 0.24 \sum X \quad (9)$$

$$X_{CA} + X_{CB} + X_{CD} = 0.26 \sum X \quad (10)$$

$$X_{DA} + X_{DB} + X_{DC} = 0.24 \sum X \quad (11)$$

2.2.3. Caso 3

O terceiro estudo de caso apresenta todo o conjunto de restrições do Caso 2. Contudo, foram considerados na função objetivo os pesos de cada fluxo OD iguais à quantidade de saltos entre pares de nós. Aos pares que possuem links diretos entre eles são atribuídos pesos iguais a 1.

$$\max X_{AB} + X_{AC} + 2X_{AD} + X_{BA} + X_{BC} + 2X_{BD} + 2X_{CA} + X_{CB} + X_{CD} + 3X_{DA} + 2X_{DB} + X_{DC}$$

3. Resultados e Discussão

Os resultados obtidos para cada caso são exibidos na Tabela 1, onde a coluna ERRO mostra a diferença percentual entre os resultados estimados e os valores reais.

Tabela 1: Erros percentuais médios e por fluxo OD para: (a)Caso 1; (b)Caso 2;(c)Caso 3

Dados Reais			Dados Estimados			ERRO			
AB	318		318			0%			
AC	289		601			108%			
AD	312		0			100%			
BA	294		882			200%			
BC	292		559			91%			
BD	267		0			100%			
CA	305		0			100%			
CB	289		1154			299%			
CD	324		903			179%			
DA	283		0			100%			
DB	277		0			100%			
DC	291		851			192%			
			Erro Médio:			130,8%			

Dados Reais			Dados Estimados			ERRO			
AB	318		281			12%			
AC	289		305			6%			
AD	312		303			3%			
BA	294		279			5%			
BC	292		259			11%			
BD	267		270			1%			
CA	305		293			4%			
CB	289		288			1%			
CD	324		304			6%			
DA	283		275			3%			
DB	277		269			3%			
DC	291		283			3%			
			Erro Médio:			4,8%			

Dados Reais			Dados Estimados			ERRO			
AB	318		75			76%			
AC	289		291			1%			
AD	312		309			1%			
BA	294		65			78%			
BC	292		271			7%			
BD	267		287			8%			
CA	305		299			2%			
CB	289		298			3%			
CD	324		78			76%			
DA	283		274			3%			
DB	277		281			2%			
DC	291		67			77%			
			Erro Médio:			27,7%			

(a)

(b)

(c)

No Caso 1, a Tabela 1a mostra um erro médio obtido de 130,8% mostrando que a utilização deste modelo de PL é inviável para aplicação prática. Vemos também que o erro máximo obtido para os fluxos individuais chega a 299%. No Caso 2, como

mostrado na Tabela 1b, a inclusão das restrições de proporcionalidade reduziu significativamente o erro médio de estimação, além de reduzir para 12% o erro máximo obtido na estimação dos fluxos individuais. Este resultado torna este modelo altamente recomendado para aplicação prática. No Caso 3, como pode ser visto na Tabela 1c, mostra que a inclusão de pesos na função objetivo alterou a solução levando a um erro médio de 27,2%, ou seja, pior que o Caso 2, porém ainda muito melhor que o obtido sem as restrições de proporcionalidade. O erro máximo para a estimação dos fluxos individuais passou de 12% para 78%.

4. Conclusões

Este trabalho propôs um novo conjunto de restrições para serem utilizadas na estimação de matrizes de tráfego utilizando PL, chamadas restrições de proporcionalidade. Os resultados anteriores apresentavam erros de estimação da ordem de 130% o que levava as conclusões de [Medina et al, 2002] pela não recomendação do método e a procura de outros métodos de estimação com aplicabilidade prática. A adição das restrições de proporcionalidade, aqui propostos, levou a resultados tão significativamente superiores que nos permitem mudar esta conclusão anterior e recomendar o novo modelo para estimação de MT. Os resultados obtidos são da mesma ordem de precisão dos mais sofisticados modelos analisados naquele trabalho. O erro médio inferior a 5% conseguido neste estudo torna o modelo de PL recomendável.

Esta é uma pesquisa em andamento. O modelo será agora testado para grandes redes de backbone e os resultados mais uma vez comparados com outros métodos. Caso a melhoria aqui obtida se confirme, o modelo de PL torna-se altamente recomendável devido a sua simplicidade e a ampla disponibilidade de software para resolvê-lo.

5. Referências

- Cao, J., Davis, D., Vander Weil, S. e Yu, B. Time-Varying Network Tomography. *Journal of the American Statistical Association*, 2000.
- Goldschmidt, O. ISP Backbone Traffic Inference Methods to Support Traffic Engineering. Em *Internet Statistic and Metrics Analysis (ISMA) Workshop*, San Diego, CA, December 2000.
- MATLAB for Windows User's Guide, The Math Works Inc., 1991
- Medina, A., Taft, N. Salamatain, K. Bhattacharyya, S., Diot, C. Traffic Matrix Estimation: Existing Techniques and New Directions. *SIGCOMM 2002*.
- Soule, A., Lakhina, A., Taft, N., Papagiannaki, K., Salamatian, K., Nucci, A., Crovella, M., e Diot, C. (2005). Traffic matrices: Balancing measurements, inference and modeling. In *Proc. of the ACM SIGMETRICS'05*, Banff, Canada.
- Teixeira, R., Duffield, N., Rexford, J., e Roughtan, M. (2005). Traffic matrix reloaded: Impact of routing changes. In *Proc. of the Passive and Active Measurement Workshop-PAM'2005*, Boston, MA, EUA.
- Tebaldi, C e West, M. Bayesian Inference of Network Traffic Using Link Count Data. *Journal of the American Statistical Association.*, pag. 557-573, june 1998.
- Zhang, Y., Roughtan, M., Duffield, N., e Greenberg, A. (2003). Fast accurate computation of large-scale IP traffic matrices from link loads. In *Proc. of the ACM SIGMETRICS'03*, San Diego, CA, EUA.